

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

Кафедра математичної інформатики



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
НЕЙРОННІ МЕРЕЖІ ТА ОСНОВИ КОМП'ЮТЕНОЇ
ЛІНГВІСТИКИ
для студентів

галузь знань **12 «Інформаційні технології»**

спеціальність **122 «Комп'ютерні науки»**

освітній рівень **бакалавр**

освітня програма **«Інформатика»**

вид дисципліни **вибіркова**

вибірковий блок **«Інтелектуальні інформаційні технології»**

Форма навчання	денна
Навчальний рік	2022/2023
Семестр	7
Кількість кредитів ECTS	4
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	іспит

Викладачі: к.ф.-м.н., асистент Тарануха В.Ю. (лекції)

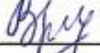
Пролонговано: на 20__/20__ н.р. _____ (_____) «__» __ 20__ р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) «__» __ 20__ р.
(підпис, ПІБ, дата)

Розробник: **Тарануха Володимир Юрійович**, к.ф.-м.н., асистент кафедри «Математичної Інформатики»
Терещенко Василь Миколайович, д.ф.-м.н., завідувач кафедри математичної інформатики

ЗАТВЕРДЖЕНО

Зав. кафедри «Математичної інформатики»

 Василь ТЕРЕЩЕНКО

Протокол № 10 від «27» 07 2024 р.

Схвалено гарантом освітньо-професійної програми «Інформатика»

«6» травня 2024 р.  Людмила ОМЕЛЬЧУК

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «6» травня 2024 року № 10

Голова науково-методичної комісії  Людмила ОМЕЛЬЧУК

ВСТУП

1. Мета дисципліни – є отримання необхідних знань з лінгвістики та комп'ютерної лінгвістики, їх подальше застосування для вирішенні різних задач з інтелектуальної обробки природномовних текстів.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

1. *Знати* дисципліни «Основи комп'ютерних алгоритмів», «Бази даних», «Дискретна математика»
2. *Вміти* виконувати інтелектуальний багатовимірний аналіз даних та їхню оперативну аналітичну обробку з візуалізацією результатів аналізу в процесі розв'язування прикладних задач у галузі комп'ютерних наук
3. *Володіти* мінімальними знаннями технічної англійської мови.

3. Анотація навчальної дисципліни:

Навчальна дисципліна «Нейронні мережі та основи комп'ютерної лінгвістики. Модуль 1. Нейронні мережі. Модуль 2. Основи комп'ютерної лінгвістики» є складовою освітньо-професійної програми підготовки фахівців за першим (бакалаврським) рівнем вищої освіти галузі знань 12 «Інформаційні технології» спеціальності 122 «Комп'ютерні науки», освітньо-професійної програми «Інформатика».

Читається на основі матеріалів «Модуль 2. Основи комп'ютерної лінгвістики»

Дана дисципліна є вибірковою навчальною дисципліною за **програмою «Інформатика»**. Викладається у 7 семестрі 4 курсу в **обсязі – 120 год.**

(4 кредити ECTS) зокрема: *лекції – 42 год., консультації – 2 год., самостійна робота – 76 год.* У курсі передбачено **3 змістових модулі, 3 модульні контрольні роботи.** Завершується дисципліна – **заліком в 7 семестрі.**

4. Завдання (навчальні цілі):

набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень у програмуванні, відповідно до кваліфікації фахівців з інформаційних технологій. Зокрема, розвивати:

- здатність оцінювати та забезпечувати якість виконуваних робіт;
- здатність проектувати та розробляти програмне забезпечення із застосуванням різних парадигм програмування: узагальненого, об'єктно-орієнтованого, функціонального, логічного, з відповідними моделями, методами й алгоритмами обчислень, структурами даних і механізмами управління;
- здатність використовувати інтелектуальні інформаційні технології.

5. Результати навчання за дисципліною:

Результат навчання (1. знати; 2. вміти; 3. комунікація)		Форми (та/або методи і технології) викладання і навчання	Методи оцінювання та пороговий критерій оцінювання (за необхідності)	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання			
PH1.1	<i>Знати основні поняття з галузі комп'ютерної лінгвістики, закони</i>	<i>Лекція</i>	<i>Тест, 55% правильних</i>	<i>15%</i>

	<i>побудови мови, базові моделі мовних явищ.</i>		<i>відповідей</i>	
<i>РН 1.2</i>	<i>Знати методи аналізу природномовних тестів та відповідні алгоритми</i>	<i>Лекція</i>	<i>Тест, 55% правильних відповідей</i>	<i>15%</i>
<i>РН 2.1</i>	<i>Вміти аналізувати мовні явища для заданої природної мови.</i>	<i>Лекція, самостійна робота</i>	<i>Тест, 55% правильних відповідей</i>	<i>30%</i>
<i>РН 2.2</i>	<i>Вміти формалізувати нестрогі методи аналізу до чітких алгоритмів та підбирати оптимальний інструментарій для виконання задачі, спираючись на машинне навчання</i>	<i>Лекція, самостійна робота</i>	<i>Тест, 55% правильних відповідей</i>	<i>30%</i>
<i>РН 3.1</i>	<i>Обґрунтовувати власний погляд на задачу, спілкуватися з колегами з питань аналізу задачі та проектування алгоритму</i>	<i>Лекція</i>	<i>Тест</i>	<i>10%</i>

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Результати навчання дисципліни	1.1	1.2	2.1	2.2	3.1
Програмні результати навчання					
ПРН19.1. Знати і застосовувати методи інтелектуального аналізу даних та штучного інтелекту, що включають методи комп'ютерної лінгвістики та комп'ютерного зору.	+	+	+	+	+
ПРН20.1. Знати методи машинного навчання для розв'язання прикладних задач, основні поняття та принципи роботи штучних нейронних мереж.				+	+

7. Схеми формування оцінки.

7.1 Форми оцінювання студентів:

- семестрове оцінювання:

1. Контрольна робота (тест): РН 1.1.— 20 балів/11 балів.

2. Контрольна робота (тест): РН 1.2.— 20 балів/11 балів.

3. Контрольна робота (тест): РН 1.2.— 20 балів/11 балів.

- підсумкове оцінювання (у формі екзамену):

- максимальна кількість балів які можуть бути отримані студентом: 40 балів;

- результати навчання які будуть оцінюватись: РН1.1, РН1.2, РН2.1, РН2.2, РН3.1.

- форма проведення і види завдань: письмова.

Види завдань: 4 теоретичних та 2 практичних завдання.

Критерії оцінювання на екзамені

Завдання	Тема завдання	Максимальний відсоток від 40 балів	Всього відсотків
Завдання 1, 2, 3, 4	Письмове запитання з наведених нижче	По 15%	75%
Завдання 5	Для заданого словника (української мови) виконати його розбір евристичним аналізом на морфемні характеристики		
Завдання 6	Для заданого набору речень англійською мовою виконати його розбір за відповідними граматиками складових та залежностей	По 25%	25%
			100%

Запитання для підготовки до екзамену

1. Закон Ціпфа.
2. Синтаксичні класи.
3. Граматичні класи.
4. Контрольована мова.
5. Регулярні граматики.
6. Контекстно-вільні граматики.
7. Трансформаційні граматики.
8. Імовірнісні контекстно-вільні граматики.
9. Граматика під'єднання дерев (Tree adjoining grammar).
10. Парсинг на основі залежностей.
11. Контекстно-вільний парсинг.
12. Алгоритм Кока-Янгера-Касамі.
13. Асоціативний підхід визначення значень багатозначних слів.
14. Автоматичне визначення значень багатозначних слів на основі статистичних даних.
15. Автоматичне визначення значень багатозначних слів на основі машинного навчання.
16. Кластеризація способів використання слів.
17. Онтології.
18. Word Net.
19. Дискурс.
20. Покрокова інтерпретація.

21. Тема і рема.
22. Тема і фокус.
23. Теорія риторичної структури (RST).
24. Теорія центрування.
25. Анафора, її типи.
26. Розв'язання анафори на основі граматичних даних.
27. Розв'язання анафори на основі семантичних даних.
28. Алгоритм послівного напіваавтоматичного поповнення словника.
29. Алгоритм групового напіваавтоматичного поповнення словника.
30. Алгоритми сегментації речень.
31. Еліптичні вхідні дані.
32. Оптимальна структура для результатів парсингу.
33. Структура Link Grammar Parser.
34. Визначення значень багатозначних слів на основі WordNet.
35. Визначення значень багатозначних слів на основі СУС.
36. Метод семантичної хмари.
37. Метод синтаксичних шаблонів.
38. Метод опроних векторів.
39. Нейронні мережі для семантичного аналізу.
40. Тема тексту. Способи представлення.
41. Визначення емоційної направленості тексту (sentiment analysis)
42. Автоматичне реферування.
43. Автоматичний переклад.

Студент не допускається до екзамену, якщо під час семестру набрав менше ніж 33 бали або менше 11 балів за будь-яку контрольну.

7.2 Організація оцінювання:

Терміни проведення форм оцінювання:

1. *Контрольна робота (тест): до 6 тижня семестру.*
2. *Контрольна робота (тест): до 11 тижня семестру.*
3. *Контрольна робота (тест): до 17 тижня семестру.*

Студент має право на одне перескладання кожної контрольної роботи із можливістю отримання максимально 80% початково визначених за цю контрольну роботу балів. Термін перескладання визначається викладачем.

У випадку відсутності студента з поважних причин відпрацювання та перездачі контрольних робіт здійснюються у відповідності до „Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу” від 1 жовтня 2010 року.

7.3 Шкала відповідності оцінок

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

8. Структура навчальної дисципліни. Тематичний план лекційних занять

№ семінару	Назва семінару	Кількість годин		
		Лекції	Семінари	Самост. робота.
Змістовий модуль 1. Явища природної мови. Морфологія та синтаксис				
1.	Тема 1. Задачі комп'ютерної лінгвістики. Природні, контрольовані та штучні мови. Закон Ціпфа. Структура слова. Слова і граматичні класи.	2		3
2.	Тема 1. Машинні словники, словники основані на корпусі текстів.	2		3
3.	Тема 2. Звичайний синтаксис. Слова і синтаксичні класи. Природні граматики та мови. Формальні граматики та мови.	2		3
4.	Тема 2. Регулярні граматики. Контекстно-вільні граматики. Трансформаційні граматики.	2		3
5.	Тема 2. Граматика під'єднання дерев (Tree adjoining grammar). Граматика залежностей.	2		4
6.	Тема 2. Парсинг на основі залежностей. Контекстно-вільний парсинг	2		4
7.	Тема 2. Алгоритм Кока-Янгера-Касамі.	1		
7.	<i>Підсумкова модульна контрольна робота 1</i>	1		4
Змістовий модуль 2. Значення слів, онтології. Дискурс. Значення тексту.				
8.	Тема 3. Онтології. Фреймові ієрархічні моделі.	2		4
9.	Тема 3. WordNet. СУС.	2		4
10.	Тема 3. Автоматичне визначення значень багатозначних слів.	2		4
11.	Тема 4. Автоматичне членування тексту. Виділення термінів.	2		4
12.	Тема 4. Дискурс. Покрокова інтерпретація. Тема і рема. Тема і фокус.	2		4
13.	Тема 4. Теорія риторичної структури (RST). Теорія центрування.	1		
13.	<i>Підсумкова модульна контрольна робота 2</i>	1		4
Змістовий модуль 3. Етапи обробки тексту.2				
14.	Тема 5. Визначення іменованих сутностей.	2		4
15.	Тема 5. Анафора(anaphora), співпосилання, її типи. Властивості. Розв'язання співпосилань (coreference resolution). Визначення значень займенників.	2		4
16.	Тема 5. Зв'язок між розв'язанням співпосилань та визначенням іменованих сутностей. Об'єднані моделі.	2		4
17.	Тема 5. Алгоритми евристичного семантичного аналізу з використанням семантичних мереж.	2		4
18.	Тема 6. Визначення теми тексту. Автоматична кластеризація потоків текстових повідомлень. Визначення спаму.	2		4
19.	Тема 6. Визначення емоційної направленості тексту (sentiment analysis)	2		4
20.	Тема 6. Автоматичне реферування.	2		4
21.	Тема 6. Автоматичний переклад на основі	1		

	імовірнісних перетворень.			
21.	<i>Підсумкова модульна контрольна робота 3</i>	1		
	ВСЬОГО	42		76

Загальний обсяг 120 год., в тому числі:

Лекції – 42 год.

Консультації - 2 год.

Самостійна робота - 76 год.

9. Рекомендовані джерела:

Основна:

1. Волошин В.Г. Комп'ютерна лінгвістика: Навчальний посібник. – Суми: Університетська книга, 2004. –382 с.
2. Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык Киев: Наук. думка, 1988.- 223 с.
3. Партико З.В. Прикладна і комп'ютерна лінгвістика, Львів, «Афіша», 2008, - 221 с.
В тому числі й інтернет ресурси
4. The Oxford handbook of computational linguistics R. Mitkov (Ed) Oxford University Press Ел. ресурс. Режим доступу: http://books.google.com.ua/books/about/The_Oxford_handbook_of_computational_lin.html?id=OaClhre-vW4C&redir_esc=y
5. Daniel Jurafsky and James H. Martin Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. 2nd edition, 2009 Ел. ресурс. Режим доступу: http://rapidlibrary.com/files/speech-and-language-processing-an-introduction-to-natural-language-processing-djv_23726564.html

Додаткова:

6. Український правопис / Ін-т мовознавства ім. О.О. Потебні НАН України, Ін-т укр. мови НАН України. — К. : Наук. думка, 2007. — 288 с.
7. The Oxford Handbook of Applied Linguistics / Ed. by R.Kaplan. NY: Oxford university press, 2002, 672 P.
8. J. Allen, “Natural language understanding” Menlo Park, Calif. Benjamin/Cummings 1995, 654 P.
В тому числі й інтернет ресурси
9. Сайт проекту WordNet Ел. ресурс. Режим доступу <http://wordnet.princeton.edu/>
10. Марчук Ю.Н. Компьютерная лингвистика М.: Изд-во Восток-Запад , 2007 г. , 317 с Ел. ресурс. Режим доступу: <http://www.twirpx.com/file/398578/>
11. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии М.: Русский мир. 2004г Ел. ресурс. Режим доступу: <http://www.twirpx.com/file/134393/>
12. Тузов В.А. Компьютерная семантика русского языка Санкт-Петербург: Издательство Санкт-Петербургского университета, 2003 Ел. ресурс. Режим доступу: <http://depositfiles.com/files/421m35w47>
13. Сайт «Автоматическая обработка текстов» Ел. ресурс. Режим доступу <http://aot.ru>
14. Український журнал комп'ютерної лінгвістики.Ел. ресурс. Режим доступу <http://franko.lviv.ua/ujcl/>
15. Спільний сайт Інституту російсько мови ім. В.В. Виноградова та компанії «СЛОВАРИ.РУ» Ел. ресурс. Режим доступу <http://www.slovari.ru/>